Google Scholar BETA

Web   Images   Video   News   Maps   **more »**

[duplicate OR replicated representative docume] [   ] - [2003] [Search]

Scholar   All articles - Recent articles   Results 1 - 10 of about 1,930 for duplicate OR replicated representative document OR webpage webcrawler OR crawler OR spider. (0.15 seconds)

Did you mean: duplicate OR replicated representative document OR **web page** webcrawler OR crawler OR spider

Method and system for detecting **duplicate** documents in web crawls - all 2 versions »
D Meyerzon, S Shoroff, FS Terek, S Norin - US Patent 6,547,829, 2003 - freepatentsonline.com
... **Representative** Image: Method and system for detecting **duplicate** ... and system for
identifying **duplicate** documents in a ... is not supported by the **document** store, the ...
Cited by 4 - Related Articles - Cached - Web Search

Engineering a multi-purpose test collection for Web retrieval experiments - all 11 versions »
P Bailey, N Craswell, D Hawking - Information Processing and Management, 2003 - Elsevier
... that CRC64 will falsely signal a **duplicate** in this ... queries for which at least one
document from the ... A **representative** distribution of server sizes was a very ...
Cited by 108 - Related Articles - Web Search

On the evolution of clusters of near-**duplicate** Web pages - all 16 versions »
D Fetterly, M Manasse, M Najork - Web Congress, 2003. Proceedings. First Latin American, 2003 - ieeexplore.ieee.org
... set of shingles to a small, yet **representative**, subset ... cluster covers all versions
of a **replicated** page ... also found that clusters of near-**duplicate** documents are ...
Cited by 42 - Related Articles - Web Search

Results from a Web Impact Factor **crawler** - all 9 versions »
M Thelwall - Journal of Documentation, 2001 - emeraldinsight.com
... common for servers to allow a **document** to be ... the pages crawled, indicating that the
duplicate pages should ... were chosen because they are **representative** of the ...
Cited by 47 - Related Articles - Web Search - BL Direct

Finding **replicated** Web collections - all 26 versions »
J Cho, N Shivakumar, H Garcia-Molina - ACM SIGMOD Record, 2000 - portal.acm.org
... The paper describes how to efficiently identify **replicated** documents and hyperlinked
document collections ... replication information to improve a **crawler** and a ...
Cited by 81 - Related Articles - Web Search - BL Direct

[PDF] Mirror, mirror on the Web: A study of host pairs with **replicated** content - all 6 versions »
K Bharat, A Broder - COMPUT. NETWORKS, 1999 - cumbrowski.com
... Host Pairs with **Replicated** Content ... that almost a third of the Web consists of **duplicate**
pages ... case the samples in the collection may not be very **representative**. ...
Cited by 81 - Related Articles - View as HTML - Web Search

Toward a Qualitative Search Engine - all 7 versions »
Y Li - 1998 - doi.ieeecomputersociety.org
... 11 we used a small but **representative** set of ... to the same **document**, detecting **duplicate**
URLs and ... search engines, information retrieval, **document** analysis, and ...
Cited by 80 - Related Articles - Web Search

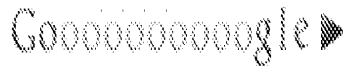[PDF] FASD: A Fault-tolerant, Adaptive, Scalable, Distributed Search Engine - all 18 versions »
AZ Kronfol - Master'sThesis, Princeton University, http://www. cs. ..., 2002 - fravia.com
... are no collisions (ie a **duplicate document** does not ... If a **document** continues to be
unpopular, even the ... most popular (and important?) data is widely **replicated**. ...
Cited by 20 - Related Articles - View as HTML - Web Search

Marie-4: A High-Recall, Self-Improving Web **Crawler** That Finds Images Using Captions - all 10 versions »
NC Rowe - 2002 - doi.ieeecomputersociety.org
... We also eliminate **duplicate** captions, and only ... candidates and picking three
representative keywords from ... the caption-likelihood and **document**-frequency factors. ...
Cited by 13 - Related Articles - Web Search - EL Direct

An efficient scheme to remove **crawler** traffic from the Internet - all 5 versions »
X Yuan, MH MacGregor, J Harms - Computer Communications and Networks, 2002. Proceedings. ..., 2002 - ieeexplore.ieee.org
... by replicating their traffic, and handing the **replicated** streams off ... out-of-order,
corrupt, and **duplicate** packets ... 10] to create a network **representative** of the ...
Cited by 3 - Related Articles - Web Search

Key authors: **A Broder** - **P Bailey** - **S Gauch** - **M Manasse** - **D Hawking**

Did you mean to search for: duplicate OR replicated representative document OR **web page** webcrawler OR crawler OR spider

Gooooooooogle ▶

Result Page: 1 2 3 4 5 6 7 8 9 10   **Next**

[duplicate OR replicated representat] [Search]

Google Home - About Google - About Google Scholar